Acta neurol. belg., 2003, **103**, 6-12

Review articles

An introduction to Bio-Inspired Artificial Neural Network Architectures

B. Fasel
IDIAP, Martigny, Switzerland

Abstract

In this introduction to artificial neural networks we attempt to give an overview of the most important types of neural networks employed in engineering and explain shortly how they operate and also how they relate to biological neural networks. The focus will mainly be on bio-inspired artificial neural network architectures and specifically to neo-perceptrons. The latter belong to the family of convolutional neural networks. Their topology is somewhat similar to the one of the human visual cortex and they are based on receptive fields that allow, in combination with sub-sampling layers, for an improved robustness with regard to local spatial distortions. We demonstrate the application of artificial neural networks to face analysis – a domain we human beings are particularly good at, yet which poses great difficulties for digital computers running deterministic software programs.

Key words : Artificial neural networks ; artificial neurons ; biological neural networks ; face detection ; facial expression recognition.

Introduction

Artificial neural networks (ANNs) are programs designed to operate functionally similar to biological nervous systems. They are based on simulated nerve cells or neurons, which are interconnected in a variety of ways to form networks that have the capacity to learn, memorize and create relationships amongst data. There are many different types of ANNs and their architecture depends on the type of task envisaged. The application of artificial neural networks is broad, however, we can distinguish a few representative categories, namely classification, forecasting and modeling. ANNs have some characteristics, which may favor them over other data analysis methods, e.g. they can deal with non-linearities of the world we live in, handle noisy or missing data and can work with a large number of variables. Artificial neural networks use highly distributed representations and transformations that operate in parallel. Therefore, ANNs are also sometimes called parallel distributed processing systems, which emphasizes the way in which the many nodes or neurons in a neural network operate in parallel.

Artificial Neural Networks are useful for a variety of applications. Originally developed as tools for the exploration and reproduction of human information processing tasks such as speech, vision, olfaction, touch, knowledge processing and motor control, they are nowadays employed for a variety of engineering tasks such as data compression, optimization, pattern recognition, system modeling, function approximation and control. For example in pattern recognition, the following tasks have been tackled : reading zip codes on envelopes (12), damage to clothes by washing powders (2), financial trading (15) and predicting suitable habitats for Tsetse flies (16). The theory that inspires neural network systems is drawn from many disciplines ; primarily from neuroscience, engineering and computer science, but also from mathematics, physics, psychology and linguistics. These sciences are working toward the common goal of building intelligent systems.

Artificial Neural Networks vs. Biological Neural Networks

Artificial Neural Networks (ANNs) are computational paradigms which implement simplified models of their biological counterparts : biological neural networks. Biological neural networks are assemblages of neurons and they share with artificial neural networks the following characteristics :

- Local information processing in neurons
- Massively parallel processing via interconnected neurons
- Acquire knowledge via learning from experience (a synapse's strength may be modified by experience)
- Information storage in distributed memory (long-term memory resides in the neurons' synapses, while short-term memory corresponds to signals passing through neurons)

Artificial neural networks learn from training data and represent a class of algorithms that allow

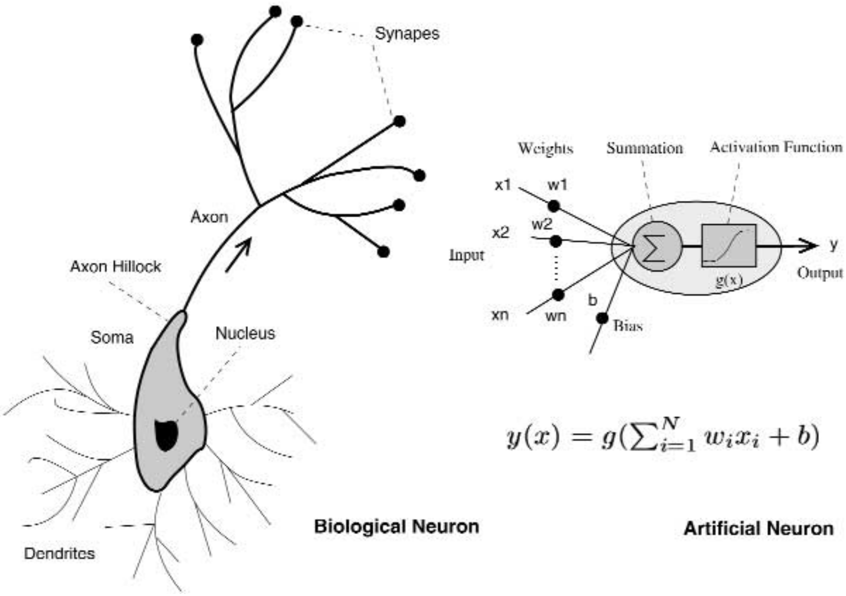

FIG. 1. — Local signal processing in neurons : shown is a representation of a biological neuron on the left hand side and a schematic of an artificial neuron on the right hand side.

for statistical modeling and prediction. The aim of neural networks is to produce a statistical models of the underlying processes from which the training data were generated in order to allow for the best possible handling of new data. Note that data processed by artificial neural networks are vectors or matrices representing any kind of signals, such as images, audio, etc. Artificial neural networks are usually determined by following properties :

- **Architecture ;** its pattern of connections between the neurons as well as the number of neurons, their respective activation functions and the number of employed layers
- **Learning Algorithm ;** its method of determining the weights on the connections

Generally speaking, we can distinguish three different types of statistical modeling problems, namely density estimation, regression and classification. In the following sections we first introduce artificial neurons and networks, before addressing network training and generalization in more detail.

Artificial Neurons

An artificial neural network consists of a large number of processing elements, called neurons. Each neuron has an internal state, called activation or activity level, which is a function of the inputs it has received. Typically, a neuron sends its activation as a signal to several other neurons. A neuron can send only one signal at a time, although that signal may be broadcasted to several other neurons. A biological neuron has three types of components that are of particular interest in understanding an artificial neuron : its dendrites, soma and axon. The many dendrites receive signals from other neurons and convey these signals via synapses towards the soma, or cell body. The soma and the enclosed nucleus don't play a significant role in the processing of incoming and outgoing data. Their primary function is to perform the continuous maintenance required to keep the neuron functional. The part of the soma that does concern itself with the signal is the axon hillock. If the aggregate input is greater than the axon hillock's threshold value, then the neuron fires, and an output signal is transmitted down the axon. The strength of the output is constant, regardless of whether the input was just above the threshold, or a hundred times as great. The output strength is unaffected by the many divisions in the axon ; it reaches each terminal button with the same intensity it had at the axon hillock. This uniformity is critical in an analogue device such as a brain, where small errors can snowball, and where error correction is more difficult than in a digital system. Each terminal button is connected to other neurons across a small gap called a synapse.

Figure 1 shows both a schematic of a biological and an artificial neuron. The latter operate from the point of view of signal processing similarly to their biological counterparts. Signals flowing into the neuron's node are modified via weights by multiplying the transmitted signal. The neuron's node sums the incoming signals and applies a threshold function, also called activation function. Under appropriate circumstances (sufficient input), the neuron transmits a single output. The output from a particular neuron may go to many other neurons (through the axon branches).

Note that artificial neurons are only similar in the way they process information, when compared to biological neurons. They are often implemented in software and represent nothing else than a mathematical function and mimic only the processing capabilities of the latter. Also note that even

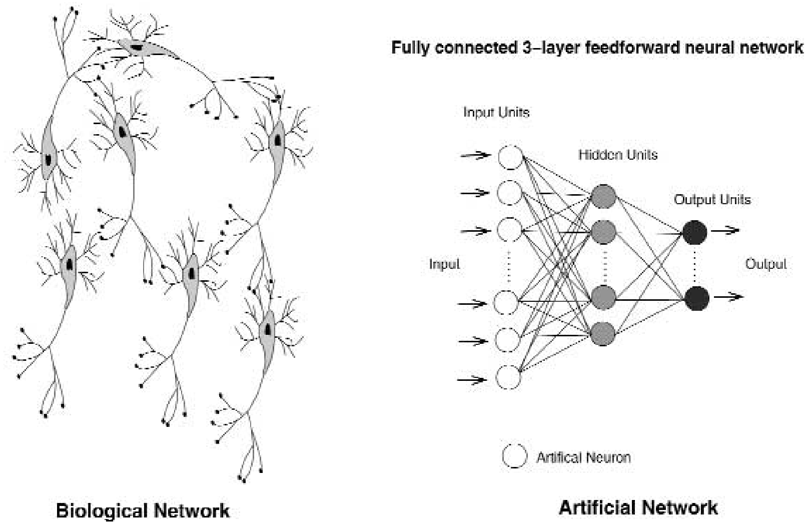

FIG. 2. — Depicted on the left hand side is a schematic of a biological neural network, while on the right hand side is shown the architecture of an artificial neural network.

though artificial neurons can be implemented in hardware and embedded in electronic artificial neural networks to operate in a truly parallel fashion, they are mostly implemented in software and run on a sequential digital computer with quantized weights. However, in the latter case, information processing is still done in parallel.

ARTIFICIAL NEURAL NETWORKS

There are many different artificial neural network architectures. No single neural network architecture is best ; rather, different architectures are useful for different applications. The most commonly used feed-forward type ANN encompass Multilayer Perceptrons (MLP) and Radial Basis Function (RBF) networks. The former operate in a global mode, where basis functions in the hidden layer cover a significant part of the input space, this in contrast to RBFs, which are referred to being local due to their respective basis functions supporting local regions in the input space. Figure 2 shows both a schematic of a biological and an artificial neural network architecture. Note that there exist more complex artificial network architectures, such as recurrent neural networks, which feature feed-back connections in addition to feed-forward connections. Feed-back connections allow them to learn context information, which is important for series predictions.

NETWORK TRAINING

In contrast to classical Artificial Intelligence (AI) approaches, ANNs with dynamic weights are not constructed using explicit rules, but statistical properties are learned from data and hyperplanes are formed that allow to separate different classes (see Fig. 3). The implicitly extracted rules are mostly not semantically accessible, even though the training methods and the network architecture, including the neuron interconnections as well as the number and type of neurons, are well known. The training algorithm plays an important role in any neural network. The latter is the process which modifies the weights and biases of the neurons, which in turn allows networks to associate certain input data patterns to certain output values. We can identify three types of learning :

- **Fixed weights :** In this case, no learning occurs (explicit rules are compiled into the network)
- **Supervised training :** Each input pattern (vector) is associated with a target output pattern (vector)
- **Unsupervised training :** No target outputs are specified (with the exception of auto-associative networks, where the targets are the same as the inputs)

Supervised training of ANNs is achieved by different algorithms that have often not very much in common with the way biological neural networks learn. These algorithms attempt to minimize the error between desired target values and output values of networks to be trained. Training samples together with desired target values are repeated in epochs until a stopping condition is met or a pre-set number of repetitions has occurred. The most commonly used algorithm for supervised training is the so called backpropagation algorithm (18), which uses the generalized delta rule. It is a gradient descent method that attempts to minimize the total squared error between the desired target value and the network output, where the computed negative of the gradient determines the direction in which the error function decreases most rapidly and thus weight updates are performed to realize this. There

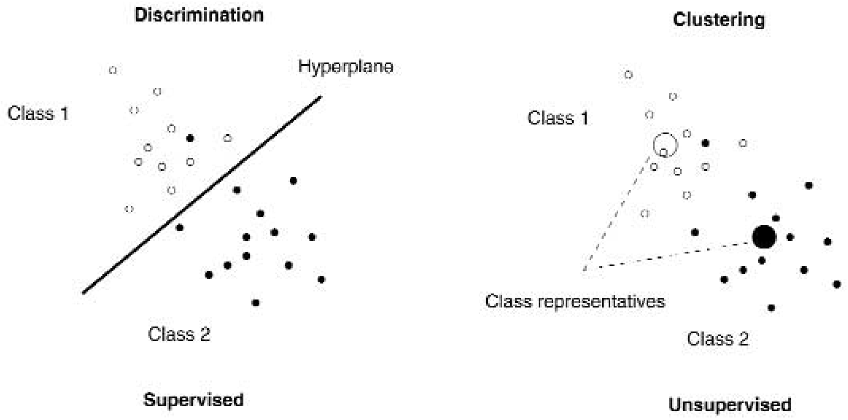

FIG. 3. — On the left hand side we can see an example for class discrimination by a hyperplane, whose position was determined during the training of the network (in this simple case being a line) and on the right hand side is given an illustration for clustering (with class representatives found using an unsupervised approach). Note that in both cases one point will be misclassified, respectively, falls into the wrong cluster.

are various algorithms for unsupervised training, the simplest and earliest rule for artificial neural networks is generally known as the Hebb rule (8). Unsupervised competitive learning is used in a wide variety of fields under a wide variety of names, the most common of which is cluster analysis, compression and data visualization.

NETWORK GENERALIZATION

One of the advantages of neural networks is their good natured degradation with regard to noisy inputs or missing data as well as their generalization performance. By training a neural network, we separate the output space into regions ; of course, the trained neural network will not only separate (classify) known input data patterns, but will also separate previously unseen data patterns. The ability of a neural network to correctly classify input data patterns that it has not seen before (has not been trained on) is termed generalization.

Bio-Inspired Neural Network Architectures

In this section we introduce a few representative ANNs that were not only inspired from biological neural networks by the way information is processed locally in nodes and transmitted over connections to other nodes, but neural network architectures that resemble also more closely their biological counter-parts with regard to the type of signals that can be measured at the output of neurons (spiking neural networks), and also with regard to the interconnections and the flow that resembles more closely to filter banks that can be found e.g. in the visual cortex.

SPIKING NEURAL NETWORKS

Most ANNs employed nowadays use analog output neurons (with rational or real output values), this in spite that biological neurons encoding information in succession of pulses and not via signal amplitude. Spiking neurons closely model biological neurons in that that they emit pulse trains (spikes) and encode information temporally. Networks based on these type of neurons are called spiking neural networks (SNNs), also called pulsed neural networks. Because of this, a SNN is well suited for applications, where the timing of input signals carries important information (e.g., speech recognition and other signal-processing applications). SNNs are capable of exploiting time as a resource for coding and computation. However, they can be applied to the same problems as non-spiking neural networks, albeit often with substantially fewer gates, see also (14).

CLUSTERING APPROACHES

Unsupervised neural network based clustering approaches allow to simulate topological sensory maps found in the brain. The self-organizing map (SOM) (11) is an unsupervised competitive network that has the ability to form topology-preserving between its input and output spaces. The defining characteristic of competitive nets is that they choose one or more output neurons that will respond to any given input pattern, instead of providing an output pattern using all output neurons. The connection weights serve as a cluster exemplar (or class representatives, see Fig. 3) instead of an input scaling function. Only the winning unit and neighboring units (in terms of the network topology, not in terms of weight vector similarity) update their weights.

Another unsupervised clustering approach is the adaptive resonance theory (ART) network (7), which has the ability to plastically adapt, when presented with new input patterns, while remaining stable at previously seen input patterns.

Convolutional Neural Networks

Neo-perceptrons (13) as well as the topologically similar Neo-Cognitrons (6), are bio-inspired hierarchical multi-layered convolutional neural network (CNN) architectures that model to some degree characteristics of the human visual cortex and encompass scale and translation invariant feature extraction layers. Neo-perceptrons can be applied directly on high-dimensional raw input images, whereby weight-sharing efficiently reduces the model complexity. Neo-perceptron are operated and trained similar to MLPs (supervised training via back-propagation algorithm). However, in contrast to the latter, they don't feature full connectivity, instead, different neuron groups extract salient features from the preceding layers or the input image. Thus, the network is forced to learn local feature extractors. Convolutional neural networks employ both simple and complex feature extractors that allow for robust object analysis with regard to spatial variations. A sample application of a neo-perceptron to facial expression recognition is described further below.

Associative Memories

Content-addressed or associative memory refers to a memory organization in which the memory is accessed by its content (as opposed to an explicit address). Thus, reference clues are associated with actual memory contents until a desirable match (or set of matches) is found. Associative memory stands as the most likely model for cognitive memories. Humans retrieve information best when it can be linked to other related information.

Associative memories can also be built with artificial neural networks. We distinguish two broad types of artificial associative memories :

– **Auto-associative Memories :** An associative memory is a system which stores mappings of specific input representations to specific output representations, or in other words, a system that associates two patterns such that when one is encountered subsequently, the other can be reliably recalled. The Hopfield model (9) is a good example for such a memory and is used as an auto-associative memory to store and recall a set of bitmap images. Given an incomplete or corrupted version of a stored image, the network is able to recall the corresponding original image.
– **Hetero-associative Memories :** Bidirectional associative memory can be viewed as a generalization of the Hopfield model and allow for a hetero-associative memory to be implemented that can e.g. the association between names and corresponding phone numbers. After training such a network, when presented with a name, it would be able to recall the corresponding phone number and vice versa. An example for a hetero-associative memory is Rosenblatt's perceptron (17).

Sample Applications : Face Analysis

Relatively simple artificial neural network architectures can cope surprisingly well with problems we human beings are good at, yet where traditional engineering approaches have difficulties. Face detection, face recognition and facial expression analysis are such tasks. In the following sections we present an automatic face detection and a facial expression recognition implementation that are based on two different artificial neuron network architectures.

Automatic Face Detection

As a first example of an artificial neural network application, we present a fully connected feed-forward multilayer perceptrons (MLP) that has been trained to detect faces in cluttered scenes (Fig. 4). The trained network is able to detect faces reliably using a mere 30 neurons in total. It detects faces by sliding a window of the size of a face through a given test image. At locations where there is a face present in the image, a value close to one can be measured at the output of the network. If there is no face present at the current location of the sliding window, the output value is close to zero. Neural networks give a probabilistic output and can miss-classify objects, thus make errors. The achieved correct recognition rate was in the range of 70-75% on a standard database, see (1) for details. The generalization performance of the network is demonstrated in the photo situated on bottom part of the right hand side of Figure 4, which shows that the network was also able to detect hand-drawn faces, even though it was trained only on photos of human faces.

Automatic Facial Expression Recognition

Another application is facial expression recognition. This task is more complicated than face detection, as there are not only two classes to be separated (faces and non-faces), but a variety of different facial expressions. We have employed convolutional neural networks for automatic facial expression analysis, a task, where we have to cope with head pose and lighting variations. Especially pose variations are difficult to tackle and many face analysis approaches presented in the literature require the use of sophisticated normalization and initialization procedures. Our data-driven face analysis approach is not only capable of extracting features relevant to the given task at hand, but is also more robust with regard to face location changes and scale variations, when compared to more traditional approaches such as e.g. multi-layer perceptrons

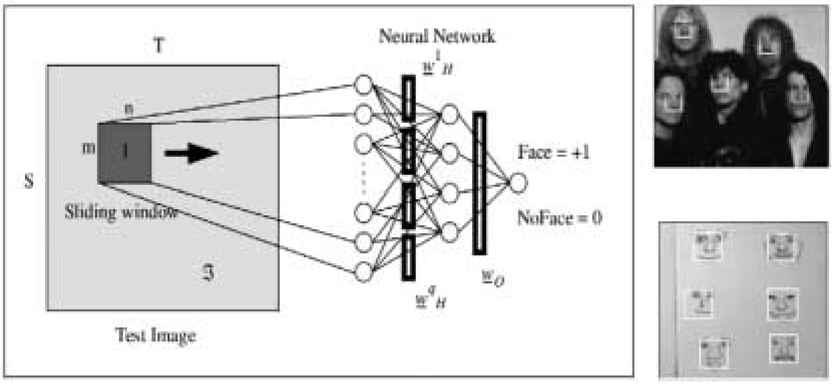

FIG. 4. — On the left hand side is shown a fully connected feed-forward multi-layer perceptron that has been trained for face detection. Depicted on the right hand side are two sample photos, where faces have been detected automatically by our network (the detected face locations are marked by white squares).

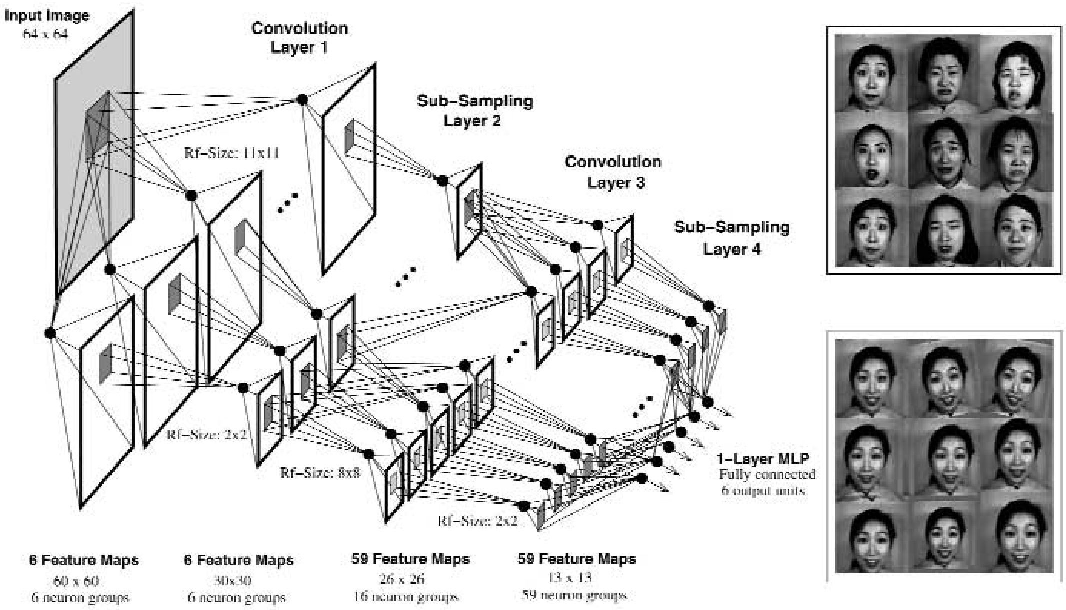

FIG. 5. — On the left hand side is depicted the architecture of a convolutional neural network (neo-perceptron) that has been trained to recognize six basic emotions, whereas on the right hand side are shown sample facial expressions of 9 subjects on top and some artificially introduced head pose variations for one subject and expression on the bottom.

(MLPs). We applied different neural network architectures, both for shape and motion recognition (4). Furthermore, we combined face identity recognition with facial expression recognition in order to obtain personalized facial expression recognition, which allowed to improve correct recognition results (5). Figure 5 shows the architecture of a convolutional neural network we employed for facial expression recognition. Note that the network architecture is composed of two parts, namely the first four layers that operate as feature extractors and the last layer that contains a fully connected MLP classifier. Hereby, the first layer of the CNN extracts simple features, while the third layer combines the inputs from the preceding layer into complex features. Layer two and four are sub-sampling layers that allow to reduce the dependency with regard to the exact location of the extracted features. The convolutional feature extractors operate like a filterbank, whose characteristics are learned from the data and which is optimally suited for the given task at hand. We obtained correct recognition rates of up to 90% for 6 basic emotions and neutral face displays on a database containing 10 female Japanese and correct recognition rates in the range of 50-80% for the same database, but with artificially increased head pose variations.

Conclusion

Artificial neural networks are powerful computer paradigms that allow to solve complex problems in engineering that would otherwise be difficult to tackle. Even very small neural networks – e.g. compare the afore mentioned neural network employed for face detection – often allow to achieve surprisingly good results.

Today, the human brain is still mostly – what engineers call – a black box, especially what concerns higher level operations encompassing reasoning and emotions. It is possible to study human behavior from the outside, as attempt psychologist and linguists or measure brain activities by using dynamic imaging techniques such as positron emission tomography (PET) or by measuring skin surface potential changes through electro-encephalography (EEG), transplanting electrodes directly into the living tissue or study the brain's architecture under a microscope.

However, all these approaches give limited insight into the mechanism of the brain and do not always allow to verify theories of how neurons communicate and complex associations occur in the brain, respectively, of how information is processed, associated, stored and represented. In this context, artificial neural networks could give valuable feedback by providing models that allow to verify theories about the brain's functioning.

Great advances have been made interfacing the brain and replace the body's sense with artificial sensors, such as artificial retinas (3, 10) and cochlea implants (19). In the future, we might also be able to directly replace certain parts of the brain by artificial implants, driven by artificial neural networks.

Acknowledgement

This work was carried out at IDIAP and was funded by the SNSF (Swiss National Science Foundation) within the framework of (IM) 2 under project number 21-54000.98. Thanks go to Alessandro Vinciarelly (IDIAP) for the fruitful discussions about artificial neural networks, which helped to improve the quality of this paper.

REFERENCES

1. BEN-YACOUB S. , FASEL B. , LUETTIN J. Fast face detection using MLP and FFT. In : *Proc. Second International Conference on Audio and Video-based Biometric Person Authentication (AVBPA'99)*, pages 31-36, 1999.
2. CARSTENSEN J. M. Description and Simulation of Visual Texture. Ph.D. thesis 59, Institute of Mathematical Statistics and Operations Research (IMSOR), Technical University of Denmark, Lyngby, 1992.
3. CHOW A. Y., PARDUE M. T., CHOW V. Y., PEYMAN G. A., LIANG C., PERLMAN J. I., PEACHEY N. S. Implantation of silicon chip microphotodiode arrays into the cat subretinal space. *IEEE Transactions on Rehabilitation Engineering*, 9 (l) : 86ff, March 2001.
4. FASEL B. Facial Expression Analysis using Shape and Motion Information Extracted by Convolutional Neural Networks. In : *International Workshop on Neural Networks for Signal Processing (NNSP 02)*, Martigny, Switzerland, 2002. IEEE.
5. FASEL B. Robust Face Analysis using Convolutional Neural Networks. In : *Proceedins of the International Conference on Pattern Recognition (ICPR 02)*, Quebec, Canada, 2002. IAPR.
6. FUKUSHIMA K. Neocognitron : A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, April 1980, **36** (4) : 193-202.
7. GROSSBERG S. Adaptive pattern classification and universal recoding : I. parallel development and coding of neural feature detectors. *Biological Cybernetics*, 1976, **23** : 121-134.
8. HEBB D. O. The Organization of Behatior : A Neuropsychological Theory. 1949.
9. HOPFIELD J. J. Neural networks and physical systems with emergent collective computation abilities. *Proceedings of the National Academy of Science*, 1982, **79** : 2554-2558.
10. HUMAYUN M. S., DE JUAN E. Jr., DAGNELIE G. Visual perception elicited by electrical stimulation of retina in blind humans. *Archives of Ophthalmology*, 1996, **114** : 40-46.
11. KOHONEN T. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 1982, **43** : 59-69.
12. LE CUN Y. , BOSER D., DENKER J. S., HENDERSON D., HOWARD R. E., HUBBARD W., JACKEL L. D. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1989, **1** : 541-551.
13. LECUN Y., BENGIO Y. Convolutional networks for images, speech, and time-series. In : ARBIB M. A. (ed.). *The Handbook of Brain Theory and Neural Networks*. MIT Press, 1995.
14. MAASS W. The Handbook of Brain Theory and Neural Networks, chapter Computation with Spiking Neurons. MIT Press (Cambridge), 2nd edition edition, 2001.
15. REFENES A.-P. *Neural Networks in the Capital Markets*. John Wiley & Sons, 1995.
16. RIPLEY B. D. Networks and Chaos – Statistical and Probabilistic Aspects, chapter Statistical aspects of neural networks, pages 40-123. Chapman & Hall, London, 1993.
17. ROSENBLATT F. Principles of Neurodynamics. Spartan, New York, 1962.
18. RUMELHART D. E., HINTON G. E., WILLIAMS R. J. Learning representations by back-propagating error. *Nature*, 1986.
19. SIMONS F. B. Electrical Stimulation of the Auditory Nerve in man. *Arch. Otolaryngol.*, 1964, **79** : 559-567.

B. FASEL,
Rue du Simplon 4,
Case Postale 592,
CH-1920 Martigny (Switzerland)
E-mail : Beat.Fasel@idiap.ch.